DOCUMENT RESUME

ED 246 083 TM 840 354

AUTHOR van der Linden, Wim J.
TITLE Choosing the Best Treatment among Different Treatment-Criterion Combinations: The Classification Problem in Educational Testing.
PUB DATE Apr 84
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bayesian Statistics; *Classification; Cutting Scores; Decision Making; Regression (Statistics); Statistical Analysis; Student Evaluation; *Student Placement; *Testing Problems
IDENTIFIERS Decision Theory

ABSTRACT

The classification problem in educational testing is a decision problem. One must assign subjects to one of several available treatments on the basis of test scores, where the success of each treatment is measured by a different criterion. Examples of classification decisions include individualized instruction, counseling, and clinical settings. Classification problems have generally been resolved using linear regression techniques. The author proposes replacing these techniques with an empirical Bayesian decision analysis. The case of classification with a threshold utility function is analyzed, and a method for finding optimal assignment rules is described. The illustration is restricted to assignment on the basis of a single test score without quota restrictions. The purpose of the paper is to contrast the differences between the traditional linear regression and a Bayesian approach to the classification problem. (Author/DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED246083

Choosing the Best Treatment among Different Treatment-Criterion Combinations: The Classification Problem in Educational Testing

Wim J. van der Linden

Twente University of Technology

Enschede, The Netherlands

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. J. van der Linden

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 840 354

Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, Louisiana, April 23-27, 1984.

Abstract

The classification problem in educational testing consists of the task to assign subjects to one out of several available treatments on the basis of test scores where the success of each of the treatments is measured by a different criterion. Examples can be found in individualized instruction, counseling, and clinical settings. To date, classification problems have been approach using lineair regression techniques. It is proposed to replace these techniques by an (empirical) Bayesian decision analysis. As an example the case of classification with a threshold utility function is analyzed and optimal assignment rules are given. Technical details and analyses of classification decisions with other types of utility functions are given in a companion paper by the same author.

Keywords: Cutting Scores, Decision Theory, Classification, Mastery Testing.

Choosing the Best Treatment among Different Treatment-Criterion Combinations: The Classification Problem in Educational Testing

One of the basic problems in test-based decision making is the classification problem. Typically, in a classification problem the decision maker is confronted with several "treatments" to which he/she has to assign subjects. For each treatment there is a qualitatively different criterion representing the success of exposing the subject to the treatment. The only thing the decision maker has at his/her disposal is the subjects' scores on a certain aptitude or performance test (battery) administered prior to the treatment.

Examples of test-based classification problems include vocational guidance situations in which most promising types of schools or careers must be identified, classification of students in individualized instructional systems with tracks at different levels, testing for the military service, selecting one out of different available therapies in clinical settings, etc.

Obviously, the classification problem is a decision problem, and as such it has definite relationships to other decision problems in educational and psychological testing. Elsewhere the author has proposed a typology of test-based decision making which, in addition to the classification decision, has the selection, mastery, and placement decisions as basic elements (van der Linden, 1984). The classification decision resembles the placement decision in that the latter is a treatment-assignment problem as well. The crucial difference, however, is that in placement decisions each treatment leads to the same criterion whereas in classification decisions each treatment has its own criterion. An example of a placement decision is the aptitude-treatment-interaction paradigm in instructional psychology where students

are allowed to reach the same educational objectives via different instructional treatments. Selection and mastery decisions differ from classification and placement decisions by the presence of only one treatment. In these decisions, it is the decision maker's task to decide whether or not to accept subjects for a certain treatment, respectively, whether or not they have profited enough from a treatment to be dismissed. The four basic types of decisions in this paragraph can be met both in their pure forms or in combinations with each other. Also, further refinements within each type are possible, for instance by imposing quota restrictions or distinguishing between subpopulations varying on a relevant attribute (cf. van der Linden, 1984).

The appropriate framework for dealing with decision problems such as the above is (empirical) Bayesian decision theory (e.g., Raiffa, 1968). Applications of Bayesian decision theory to the selection, mastery and placement decision problems are amply available now (e.g., Chuang, Chen & Novick, 1981; Cronbach & Gleser, 1965; Huynh, 1976; van der Linden, 1980, 1981). To date, however, the classification problem has been devoid of decision-theoretic analysis. The traditional approach to this type of problem has been the use of linear-regression techniques in which for each treatment the regression line of its criterion score on the test score is estimated and subjects are assigned to treatments on the basis of their predicted criterion scores. In the case of a test battery instead of a single test, multiple-regression analysis has been the choice. This approach has led to the concept of differential validity (Horst, 1955; Mollenkopf, 1950) which has had a prominent place in the literature on constructing tests for classification decisions. However, in spite of its long tradition, it is not difficult to show why a linear-regression approach to the classification problem is inferior to (empirical) Bayesian decision analysis. An obvious argument is the supposed linearity of the regression functions for each of the treatments.

But more important is the fact that the linear-regression approach ignores the utility structure of the decision problem and offers no decision rules that are optimal by obvious standards. Also, decision analysis is a more general approach in the sense that for certain specific assumptions concerning probability distributions and utility functions it can be shown to yield the linear regression solution.

The purpose of this paper is to demonstrate how the classification problem can be formalized as a problem of Bayesian decision making. In particular the case of classification with a threshold utility function is analyzed, and it is indicated how optimal assignment rules can be found. The attention will be restricted to assignment on the basis of a single test score and without quota restrictions. A more detailed treatment of the classification problem, dealing with such restrictions as well as with other classes of utility functions, is given in a companion paper by the same author, which is available on request.

Classification Decisions with Threshold Utility

Suppose a series of subjects, who can be considered as randomly drawn from some population P, must be classified into $t + 1$ treatments indexed by $j = 0, 1, \ldots, t$. Each treatment leads to a certain distribution of performances of P on its associated criterion which is denoted by a random variable Y_j with range R_j. Y_j will be considered to be continucus (although in many applications Y_j may be discrete). The test scores observed prior to the treatment are denoted by a random variable X with discrete values $x = 0, \ldots, n$. It is assumed that, for X and Y_j, P yields a joint distribution with probability (density) function $\eta_j(x,y_j)$.

In the present paper, the classification problem is supposed to have a monotone form. This restricts the applicability of the results somewhat since it imposes certain conditions on the probability distributions and the utility structure. On the other hand, it simplifies the results because the optimal decision rules looked for now take the form of a series of cutting scores on the test

$$(1) \qquad 0 = c_0 \leq c_1 \leq \ldots \leq c_j \leq \ldots \leq c_{t+1} = n \qquad \text{with } t \leq n,$$

such that treatment j is assigned in the event of $c_j \leq X < c_{j+1}$ (for $j = t$ the second inequality is not strict). For the probability distributions involved, the monotony assumption requires that the distributions of Y_j given $X = x$ are stochastically increasing (cf. Chuang, Chen & Novick, 1981). The monotony assumption also imposes a condition on the utility structure that follows from Karlin and Rubin (1956). For this presentation it is enough to observe that this condition implies that the utility structure orders the treatments from "low" to "high". As an example, the case of selecting an appropriate school in guidance situations will be considered where each school teaches globally the same subjects but at different levels and schools at higher levels represent a larger utility value. Throughout this paper it is assumed that the treatments are in appropriate order reflected by their index j.

Suppose each of the schools is finished with a certifying exam which the student may pass or fail. Formally, the classification problem can then be represented as a problem with threshold utility. An appropriate threshold utility function is

$$(2) \qquad u_j(y_j) = \begin{cases} w_j & \text{for } y_j \geq d_j \\ v_j & \text{for } y_j < d_j, \end{cases}$$

where y_j is now interpreted as, e.g., the grade-point average, d_j is the minimal value of y_j for which a student passes, and w_j and v_j are the utility of finishing school j succesfully and not succesfully, respectively.

With the aid of such utility functions as the above, optimal cutting scores can be found. For each possible series of values for $(c_1, \ldots, c_n)$ the expected utility of the decisions for P can be calculated as

$$(3) \qquad B(c_1,\ldots,c_t) \equiv \sum_{j=0}^{t} \sum_{x=c_j}^{c_{j+1}-1} \int_{R_j} u_j(y_j)\eta_j(x,y_j)dy_j \ .$$

The set of optimal cutting scores in the Bayesian sense is the choice of values for $(c_1, \ldots, c_t)$ maximizing (3).

Case of t = 1

First the case of two treatments is considered and then the results will be generalized to t > 1.

Substituting (2) into (3) for t = 1 and omitting the index of c gives

$$(4) \qquad B(c) = \sum_{x=0}^{c-1} \int^{d_0} v_0\eta_0(x,y_0)dy_0 + \sum_{x=0}^{c-1} \int_{d_0} w_0\eta_0(x,y_0)dy_0 +$$

$$\sum_{x=c}^{n} \int^{d_1} v_1\eta_1(x,y_1)dy_1 + \sum_{x=c}^{n} \int_{d_1} w_1\eta_1(x,y_1)dy_1$$

which, on completing the first two sums, can be written as

$$(5)\qquad B(c) = v_0 - v_0 \sum_{x=c}^{n} \lambda_0(x) + (w_0-v_0) \sum_{x=0}^{n} \int_{d_0} \eta_0(x,y_0)dy_0 -$$

$$(w_0-v_0) \sum_{x=c}^{n} \int_{d_0} \eta_0(x,y_0)dy_0 + v_1 \sum_{x=c}^{n} \lambda_1(x) +$$

$$(w_1-v_1) \sum_{x=c}^{n} \int_{d_1} \eta_1(x,y_1)dy_1.$$

where $\lambda_j(x)$ is the marginal probability function of X.

As

$$\lambda_j(x) = \lambda(x) \qquad \text{for all values of } j,$$

and denoting the probability distribution function of Y_j given $X = x$ by $\Omega_j(y_j|x)$, it follows that (5) can be reduced to

$$(6)\qquad B(c) = \text{constant} + \sum_{x=c}^{n} \left\{(w_0-v_0)\Omega_0(d_0|x)-(w_1-v_1)\Omega_1(d_1|x)\right\} \lambda(x)$$

Since $\lambda(x)$ is nonnegative for all values of x, it holds that (6) is maximal for the smallest value of x for which

$$(7)\qquad (w_0-v_0)\Omega_0(d_0|x) - (w_1-v_1)\Omega_1(d_1|x)$$

is positive. The (strict) monotony of the problem assumed here guarantees that this solution exists and is unique.

It is important to note that this solution depends only on the difference between w_j and v_j and not on their individual values. This simplifies the decision maker's task of specifying (2) to some extent (see below). In practical situations, the solution is found by estimating the conditional distributions of criterion scores given predictor scores, checking for stochastical increasingness, and substituting the estimates. If not enough data are

available, it is also possible to fit a model for the distributions and estimate the solution under this model.

An interesting case arises if it can be assumed that (approximately) $w_0 - v_0 = w_1 - v_1$. Then (7) reduces further to

$$(8) \qquad \Omega_0(d_0|x) - \Omega_1(d_1|x) \ ,$$

and it is no longer needed to specify even the values of $w_j - v_j$. If normal distributions of Y_j given $X = x$ can be assumed, (8) is replaced by

$$\frac{d_0 - E_0(Y|x)}{[Var_0(Y|x)]^{\frac{1}{2}}} - \frac{d_1 - E_1(Y|x)}{[Var_1(Y|x)]^{\frac{1}{2}}}$$

where $E_j(Y|x)$ and $Var_j(Y|x)$ are the conditional expectations and variances, respectively. Further approximations are now possible. For example, if it holds (approximately) that $Var_0(Y|x) = Var_1(Y|x)$ for relevant values of x, then the optimal cutting score between the two treatments is the smallest value of x for which

$$(9) \qquad E_1(Y|x) - E_0(Y|x) > d_1 - d_0$$

is positive. Also, with linear regression lines $E_j(Y|x) = \alpha_j + \beta_j x$ and homoscedasticity $Var_j(Y|x) = Var_j(Y.x)$. The optimal cutting score c^* is now equal to the smallest integer value larger than or equal to

$$(10) \qquad \frac{(d_1-\alpha_1)[\mathrm{Var}_0(Y.x)]^{\frac{1}{2}} - (d_0-\alpha_0)[\mathrm{Var}_1(Y.x)]^{\frac{1}{2}}}{\beta_1[\mathrm{Var}_0(Y.x)]^{\frac{1}{2}} - \beta_0[\mathrm{Var}_1(Y.x)]^{\frac{1}{2}}}$$

To calculate this value only estimates of the regression parameters and the pooled variances of Y given X = x are required.

Case of t > 1

The generalization to more than 2 treatments with threshold utility follows from a derivation along the same line as in (4) - (8). The results turn out to be simple and boil down to applying the above findings to each pair of adjacent treatments. This procedure parallels the one for placement decisions with more than 2 treatments (van der Linden, 1981).

Concluding Remarks

Two further comments are necessary. First, it is noted that the decision maker's task of specifying his/her utility parameter values, usually a delicate matter, is a relatively simple one in the present case of a threshold utility function. The decision maker only needs to provide values for the difference $w_j - v_j$. For the example adopted in this paper this amounts to specifying for each alternative school the difference in utility between passing and failing its final exam. This task can be performed using a lottery method well-known in this area (e.g. Luce & Raiffa, 1957, chap. 2) or via direct scaling of preferences (Vrijhof, Meilenbergh & van den Brink, 1983). If it is believed that the final exam is not the ultimate criterion but that this lies in its civil effects,

psychological well being, etc., Keeney and Raiffa's handling of proxy attributes may be useful (Keeney & Raiffa, 1976, sect. 2.5). An interesting case has arosen for $w_j - v_j = w - v$ for all values of j. In our example, if the difference in utility between passing and failing is the same for each school, the solution under this condition reduces to the comparison in (8) and no specifications of the threshold utility parameters are needed at all.

Note that (10), although its distributional assumptions contain the regression lines of criterion on predictor scores, still does not coincide with the results of the traditional linear regression approach. The difference is the presence of the thresholds d_j and the pooled variances $Var_j(Y.x)$ in (10). As soon as d_j and $Var_j(Y.x)$ are treatment-independent constants, (10) reduces further to $(\alpha_0-\alpha_1) / (\beta_1-\beta_0)$ which is the value of x for which the regression lines cross and on which the traditional approach is based. The traditional solution can thus be conceived of as a Bayes rule under threshold utility (2) with equal differences between its parameter values for all treatments, homoscedasticity of the variances of Y given $X = x$ in all treatments, and equality of these conditional variances across all treatments.

It was the purpose of this paper to give the reader the flavor of such differences between the traditional linear regression and a Bayesian approach of the classification problem.

References

Chuang, D.T., Chen, J.J. & Novick, M.R. Theory and practice for the use of cut-scores for personnel decisions. Journal of Educational Statistics, 1981, 6, 129-152.

Cronbach, L.J. & Gleser, G.L. Psychological tests and personnel decisions (2nd ed.). Urbana, Ill.: University of Illinois Press, 1965.

Horst, P. A technique for the development of a multiple absolute prediction battery. Psychological Monograph, 1955, No. 390.

Huynh, H. Statistical considerations of mastery scores. Psychometrika, 1976, 41, 65-79.

Karlin, S. & Rubin, H. The theory of decision procedures for distributions with monotone likelihood ration. Annals of Mathematical Statistics, 1956, 27, 272-299.

Keeney, R.L. & Raiffa, H. Decisions with multiple objectives: Preferences and value trade offs. New York: Wiley, 1976.

Luce, R.D. & Raiffa, H. Games and decisions. New York: John Wiley & Sons, Inc., 1957.

Mollenkopf, W.G. Predicted differences and differences between predictions. Psychometrika, 1950, 15, 409-417.

Raiffa, H. Decision analysis. Reading, Mass.: Addison-Wesley, 1968.

van der Linden, W.J. Decision models for use with criterion-referenced tests. Applied Psychological Measurement, 1980, 4, 469-492.

van der Linden, W.J. Using aptitude measurements for the optimal assignment of subjects to treatments with and without mastery scores. Psychometrika, 1981, 46, 257-274.

van der Linden, W.J. Decision theory in educational research and testing. In T. Husen & T.N. Postlethwaite, International encyclopedia of education:

Research and studies. Oxford: Pergamon Press, 1984. (In press.)

Vrijhof, B.J., Mellenbergh, G.J. & van den Brink, W.P. Assessing and studying utility functions in psychometric decision theory. Applied Psychological Measurement, 1983, 7, 341-357.

Author's Address

Wim J. van der Linden, Afdeling Toegepaste Onderwijskunde, Technische Hogeschool Twente, Postbus 217, 7500 AE Enschede, The Netherlands.